

STATISTICAL REVIEW AND EVALUATION **AMENDMENT**

NDA #:	21-272
Applicant:	United Therapeutics Corporation
Name of Drug:	Remodulin™, formerly Uniprost™ (treprostinol sodium)
Indication:	Treatment for pulmonary arterial hypertension
Document reviewed:	Amendment with electronic data set
Date of submission:	April 12, 2001
Statistical Reviewer:	John Lawrence, Ph.D. (HFD–710)
Medical Reviewer:	Abraham Karkowsky, M.D. (HFD–110)

On April 11, 2001, the sponsor met with the FDA to discuss additional analyses correlating the primary endpoint of exercise tolerance and Borg Dyspnea score in the two pivotal trials. Subsequently, an amendment was filed by the sponsor to include these additional analyses.

The primary analysis of exercise tolerance used the standardized midrank of the residuals from a linear regression model. In certain instances, a worst score was imputed. After carrying forward last observations for missing values at Week 12 and the worst score imputation, the standardized midranks were then re-calculated. These are numbers between 0 and 1 which measure how much a patient improved their walking distance from baseline, after adjusting for particular covariates.

The Borg Dyspnea score is a number between 0 and 10 that is reported by the patient at the completion of the 6-minute walk test (0=no dyspnea to 10=maximum dyspnea). Changes from baseline in Borg scores at each visit were calculated. The standardized midranks were calculated and the last observation was carried forward for missing values at Week 12 (worst scores were imputed for treatment failures). Standardized midranks were then re-calculated. Since low scores show an improvement, the standardized midranks were inverted (subtracted from 1), to make high ranks associated with improvement.

For each patient, the Walk ranks and the Borg ranks were averaged together in an attempt to create a variable that measures simultaneously how much further a patient could walk and how much effort it took the patient to walk.

Figure 1a shows the boxplots of the Walk ranks for the placebo and treatment group. This figure shows that there is a numerical difference in the median scores for the

two groups, favoring the active treatment group. The primary analysis also shows this because the p-value is 0.0064 [sponsor's analysis]. Figure 1b shows the boxplots of Borg ranks for the two groups. The difference here is apparently more significant ($p=0.000103$). Figure 1c shows the boxplots of the standardized midranks derived by averaging the Walk ranks and Borg ranks. Figures 1b and 1c look nearly identical and the p-values are also very close. A possible explanation is that when two variables favor the same group and those two variables are combined together in some way, the difference between the two groups with respect to the combined endpoint might be expected to be even more significant than the difference with respect to either one alone (see appendix). In summary, the apparent difference between the two groups in the combined endpoint might be almost completely explained by the difference between the two groups in Borg ranks alone. Therefore, it might be more useful to look at this variable alone rather than trying to combine the two variables together into something that is difficult to interpret.

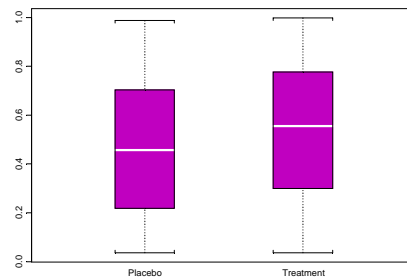


Figure 1a. Boxplots of Walk ranks alone ($p=0.0064$)

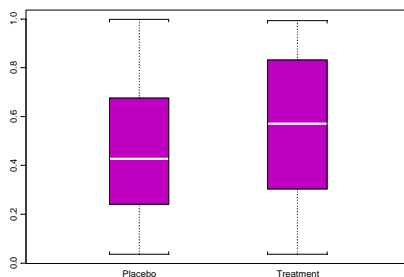


Figure 1b. Boxplots of Borg ranks alone ($p=0.000103$)

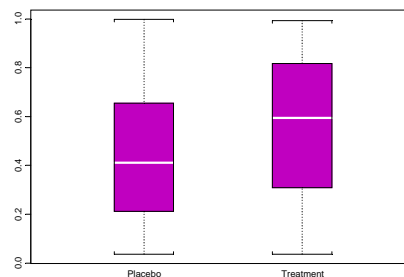


Figure 1c. Boxplots of Borg+Walk ranks ($p=0.0000084$)

410 out of the 470 patients randomized had an exercise test at Week 12. The patients in the treatment group showed a median improvement of 1 point from baseline and the patients in the placebo group showed virtually no change from baseline (median

change of 0). A summary of the changes in Borg score from baseline appears in Table 1. To make this table, fractions resulting from averaging two baseline values are rounded up. Randomization was stratified within three subgroups: primary pulmonary hypertension (PPH), secondary pulmonary hypertension with vasodilators (SPHV), and secondary pulmonary hypertension without vasodilators (SPHnoV). Among the 40 patients in the treatment group that improved by 3 or more points, 24 were in the PPH subgroup, 4 were in the SPHV subgroup, and 12 were in the SPHnoV subgroup. Among the 17 patients in the placebo group that improved by 3 or more points, 8 were in the PPH subgroup, 3 were in the SPHV subgroup, and 6 were in the SPHnoV subgroup. These particular factors do not clearly indicate that one type of patient is more likely to receive a benefit (measured by an improvement in Borg score by at least 3 points) than another.

Table 1. Summary of changes in Borg Dyspnea scores from baseline among patients with a 12-Week exercise test.

	Improved from baseline by			No change	Deteriorated from baseline by		
	3 + points	2 points	1 point		1 point	2 points	3 + points
Treatment	40	33	51	45	18	7	6
Placebo	17	22	46	69	25	15	16

In conclusion, this reviewer recommends that any weight given to the changes in Borg score should stand on the merits of the Borg score variable itself without combining this variable together with the primary variable. There are different ways to combine the two variables and the combined endpoint may be difficult to interpret. Even if there was agreement on the correct way to combine them, the p-value from the analysis of the combined endpoint seems to be precipitated by the Borg scores alone. The Borg score itself is not an objective "hard" endpoint. Since it is reported by the patient, it is subject to unconscious bias.

Appendix

Suppose $\begin{pmatrix} X_1 \\ Y_1 \end{pmatrix}, \begin{pmatrix} X_2 \\ Y_2 \end{pmatrix}, \dots, \begin{pmatrix} X_n \\ Y_n \end{pmatrix}$ are iid with mean $\begin{pmatrix} 0 \\ 0 \end{pmatrix}$ and covariance $\begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}$ under the null hypothesis. Here, the first component is the Walk rank and the second component is the Borg rank for the patients in the treatment group (standardized to have mean 0 and variance 1 under the null hypothesis). The standardized statistics using each variable separately are $\frac{\sum X_i}{\sqrt{n}}$ and $\frac{\sum Y_i}{\sqrt{n}}$. Actually, the statistics used here are stratified and the variables are not independent, but the argument is easier to understand in this simpler scenario. The statistic using the combined endpoint derived by adding the two components (making the argument easier again by assuming that midranks are not re-

calculated) is $\frac{\sum X_i + Y_i}{\sqrt{n + n + 2n\rho}} = \sqrt{\frac{2}{1 + \rho}} \frac{\frac{\sum X_i}{\sqrt{n}} + \frac{\sum Y_i}{\sqrt{n}}}{2}$. By writing the statistic this way, we see that the statistic derived from this combined endpoint is the average of the two individual statistics, multiplied by the constant $\sqrt{\frac{2}{1 + \rho}}$. This constant is always greater than 1. For example, if the two individual statistics are close to each other and the correlation is about 0, then the combined statistic will be about $\sqrt{2} \approx 1.41$ times as large as the common value, or roughly 40% larger. In this case, the correlation between the two variables is about 0.4 and the statistics from the individual variables are approximately 2.7 and 4.4. From this argument, we would expect the combined to statistic to be roughly $\sqrt{\frac{2}{1.4}} \frac{2.7 + 4.4}{2} \approx 4.24$ (in fact, it is 4.45). The message from this argument is that when the p-value from one variable is very small and this variable is combined with a second variable, it should not be surprising that the p-value from the sum of the two variables is very small.

John Lawrence, Ph.D.
Mathematical Statistician

This review consists of 5 pages of text, tables, and figures.

Concur: James Hung, Ph.D.
Acting Team Leader, Biometrics I

George Chi, Ph.D.
Division Director, Biometrics I

cc: NDA # 21-272
HFD-110/Dr. Lipicky
HFD-110/Dr. Karkowsky
HFD-110/Mr. Fromme
HFD-700/Dr. Anello
HFD-710/Dr. Chi
HFD-710/Dr. Hung
HFD-710/chron

LAWRENCEJ/594-5375/amend4172001.doc/07/18/01